Name : Dawood Sarfraz

Rollno : 20P-0153 ¶

Task no : 02

In [1]:
```
ls
```

```
'Pandas Tasks.ipynb'   titanic.csv
```

In [2]:
```
import warnings
warnings.filterwarnings("ignore")
```

In [3]:
```
import pandas as pd
```

In [4]:
```
df = pd.read_csv("titanic.csv")
```

In [5]:
```
df
```

Out[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7

891 rows × 12 columns

In [6]:
```
df.head(10)
```

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25(
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.28:
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.92:
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.10(
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05(
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.45:
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.86:
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.07:
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.13:
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.07(

In [7]:
```
df.tail(10)
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7

In [8]:
```
df.sample(10)
```

Out[8]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
548	549	0	3	Goldsmith, Mr. Frank John	male	33.0	1	1	363291	20.525
680	681	0	3	Peters, Miss. Katie	female	NaN	0	0	330935	8.137
315	316	1	3	Nilsson, Miss. Helmina Josefina	female	26.0	0	0	347470	7.854
681	682	1	1	Hassab, Mr. Hammad	male	27.0	0	0	PC 17572	76.729
475	476	0	1	Clifford, Mr. George Quincy	male	NaN	0	0	110465	52.000
51	52	0	3	Nosworthy, Mr. Richard Cater	male	21.0	0	0	A/4. 39886	7.800
231	232	0	3	Larsson, Mr. Bengt Edvin	male	29.0	0	0	347067	7.775
699	700	0	3	Humblen, Mr. Adolf Mathias Nicolai Olsen	male	42.0	0	0	348121	7.650
76	77	0	3	Staneff, Mr. Ivan	male	NaN	0	0	349208	7.895
209	210	1	1	Blank, Mr. Henry	male	40.0	0	0	112277	31.000

In [9]:
```
# Q-1
rows, columns = df.shape
print("Rows = ", rows)
print("columns = ", columns)
```

```
Rows =  891
columns =  12
```

In [10]:
```
df["Age"]
```

Out[10]:
```
0      22.0
1      38.0
2      26.0
3      35.0
4      35.0
       ...
886    27.0
887    19.0
888     NaN
889    26.0
890    32.0
Name: Age, Length: 891, dtype: float64
```

In [11]:
```
# Q-2
average_age = df['Age'].mean()
print("Average Age = ", round(average_age, 3))
```

```
Average Age =  29.699
```

In [12]:
```
df["Survived"] # Im assumming 0 for Not survived and 1 for survived
```

Out[12]:
```
0      0
1      1
2      1
3      1
4      0
      ..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

In [13]:
```
# Q-3
survived = df['Survived'].sum()
not_survived = len(df) - survived
print("Survived = ", survived)
print("Not Survived = ", not_survived)
```

```
Survived =  342
Not Survived =  549
```

In [14]:
```
df["Sex"]
```

Out[14]:
```
0        male
1      female
2      female
3      female
4        male
        ...
886      male
887    female
888    female
889      male
890      male
Name: Sex, Length: 891, dtype: object
```

In [15]:
```
# Q-4
genders = df['Sex'].value_counts()
total = len(df)
males = (genders['male'] / total) * 100
females = (genders['female'] / total) * 100
print("Male Percentage = ", males)
print("Female Percentage = ", females)
```

```
Male Percentage =  64.75869809203144
Female Percentage =  35.24130190796858
```

In [16]:
```
genders
```

Out[16]:
```
Sex
male      577
female    314
Name: count, dtype: int64
```

In [38]:
```
# Q-5
average_age_by_class = df.groupby('Pclass')['Age'].mean()
print("Average Age by Passenger Class = ", average_age_by_class)
```

```
Average Age by Passenger Class =  Pclass
1    38.233441
2    29.877630
3    25.140620
Name: Age, dtype: float64
```

In [18]:
```
df["Pclass"]
```

Out[18]:
```
0      3
1      1
2      3
3      1
4      3
      ..
886    2
887    1
888    3
889    1
890    3
Name: Pclass, Length: 891, dtype: int64
```

In [40]:
```
# Q-6
average_fare_by_class = df.groupby('Pclass')['Fare'].mean()
print("Average Fare by Passenger Class = ", average_fare_by_class)
```

```
Average Fare by Passenger Class =  Pclass
1    84.154687
2    20.662183
3    13.675550
Name: Fare, dtype: float64
```

In [20]:
```
# Q-7
survived_by_port = df[df['Survived'] == 1]['Embarked'].value_counts
print("Survived by Embarkation Port = ", survived_by_port)
```

```
Survived by Embarkation Port =  Embarked
S    217
C     93
Q     30
Name: count, dtype: int64
```

In [21]:
```
# Q-8
oldest = df[df['Age'] == df['Age'].max()]
youngest = df[df['Age'] == df['Age'].min()]
print("Oldest Passenger:")
print(oldest[['Name', 'Age']],"\n")
print("Youngest Passenger:")
print(youngest[['Name', 'Age']])
```

```
Oldest Passenger:
                                        Name   Age
630  Barkworth, Mr. Algernon Henry Wilson  80.0

Youngest Passenger:
                                    Name   Age
803  Thomas, Master. Assad Alexander  0.42
```

In [22]:
```
# Q-9
siblings_spouses = len(df[df['SibSp'] > 0])
parents_children = len(df[df['Parch'] > 0])
```

In [41]:
```
print("Sibling Spouses = " , siblings_spouses)
```

```
Sibling Spouses =  283
```

In [42]:
```
print("Parents Spouses = " ,parents_children)
```

```
Parents Spouses =  213
```

In [43]:
```
# Q-10
age_bins = [0, 18, 65, 100]
age_labels = ['Child', 'Adult', 'Elderly']
df['AgeCategory'] = pd.cut(df['Age'], bins=age_bins, labels=age_lab
survival_rate_age = df.groupby('AgeCategory')['Survived'].mean()
```

In [45]:
```
print("Survival Rate by Age = ",survival_rate_age)
```

```
Survival Rate by Age =  AgeCategory
Child      0.503597
Adult      0.386243
Elderly    0.125000
Name: Survived, dtype: float64
```

In [46]:
```
# Question 11
alone_passengers = len(df[(df['SibSp'] == 0) & (df['Parch'] == 0)])
with_family_passengers = len(df[(df['SibSp'] > 0) | (df['Parch'] > 
survival_rate_alone = df[df['SibSp'] == 0]['Survived'].mean()
survival_rate_with_family = df[(df['SibSp'] > 0) | (df['Parch'] > 0
print("Alone Passengers = ", alone_passengers)
print("With Family Passengers = ", with_family_passengers)
print("Survival Rate Alone Passengers = ", survival_rate_alone)
print("Survival Rate with Family Passengers = ", survival_rate_with_
```

```
Alone Passengers =  537
With Family Passengers =  354
Survival Rate Alone Passengers =  0.34539473684210525
Survival Rate with Family Passengers =  0.5056497175141242
```

In [47]:
```
# Question 12
survival_rate_by_embarkation = df.groupby(['Embarked', 'Pclass'])['
```

In [48]:
```
print("Survival Rate by Embarkation = ", survival_rate_by_embarkati
```

```
Survival Rate by Embarkation =  Embarked  Pclass
C         1         0.694118
          2         0.529412
          3         0.378788
Q         1         0.500000
          2         0.666667
          3         0.375000
S         1         0.582677
          2         0.463415
          3         0.189802
Name: Survived, dtype: float64
```

In [49]:
```
# Question 13
correlation_age_fare = df['Age'].corr(df['Fare'])
```

In [50]:
```
print("Co-relation Age Fare = ", correlation_age_fare)
```

```
Co-relation Age Fare =  0.09606669176903887
```

In [51]:
```
# Question 14
most_common_embarkation_port = df['Embarked'].mode()[0]
```

In [52]:
```
print("Most Common Embarktion Port = ", most_common_embarkation_por
```

```
Most Common Embarktion Port =  S
```

In [53]:
```
# Question 15
passengers_with_more_than_one_sibling_spouse = len(df[df['SibSp'] >
```

In [54]:
```
print("Passengers with more than one sibling or spouse aboard:", pa
```

```
Passengers with more than one sibling or spouse aboard: 74
```

In [55]:
```
# Question 16
passengers_with_family_by_class = df[(df['SibSp'] > 0) | (df['Parch
```

In [56]:
```
print("Passengers with family members by class = ", passengers_with
```

```
Passengers with family members by class =  Pclass
1    107
2     80
3    167
Name: PassengerId, dtype: int64
```

END of TASK

In [ ]:
```

```